Analyze and Interpret Surveillance Data

Created: 2013
Table of Contents

ANALYZE AND INTERPRET SURVEILLANCE DATA ................. 3

LEARNING OBJECTIVES ..................................................................................................... 3
ESTIMATED COMPLETION TIME ...................................................................................... 3
REFERENCES AND RESOURCES ...................................................................................... 3
MODULE CONTENT ............................................................................................................ 4
SKILL ASSESSMENT ......................................................................................................... 16
Analyze and Interpret Surveillance Data

Learning Objectives

At the end of the training, you will be able to:

• Describe data to collect based on the objective of a surveillance system.
• Identify how to present surveillance data.
• Interpret surveillance data, including trends and patterns.
• Identify groups and geographic areas to be targeted for interventions, based on surveillance data.

Estimated Completion Time

• 3 hours, 30 minutes (1 hour, 50 minutes interactive lecture; 1 hour 40 minutes skill assessment)

References and Resources

• Central America FETP Basic and Intermediate Curriculum
• Conducting Surveillance
• Organizing and Presenting
• Surveillance Interpretation and Analysis
• WHO Global Infobase. https://apps.who.int/infobase/Comparisons.aspx
• WHO Non-communicable Disease Profile, United Republic of Tanzania. https://apps.who.int/infobase/CountryProfiles.aspx
### Lesson Overview

Review of data management
- Categories of data
- Confidentiality
- Data quality

Analyze and interpret data
- Presentation
- Descriptive epidemiology
- Risk factors
- Limitations when interpreting data

### REVIEW OF DATA MANAGEMENT

### Categories of Data

- Identifying
- Demographic
- Clinical
- Laboratory
- Risk factor
- Source

**Question:** What kinds of information would you find in each of these categories of data?
**Confidentiality**

- All data must be kept confidential.
- Some surveillance systems do not enter identifying information to protect confidentiality.

**Maintaining Confidentiality**

**Question:** Why is it important to keep surveillance data confidential?

**Ways to protect confidentiality:**
- Assign a unique ID number to each case or record
- Avoid unintentional disclosure

**Data Quality: Completeness**

Completeness can refer to many things:
- **Completeness of data collected**
  - How much of the data are missing?
- **Completeness of reporting**
  - Did the surveillance system capture all of the events?

**Key Point:** Data should be of high quality and complete.
**Question: Completeness**

Table 1. Diabetes

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>491</td>
<td>49.1</td>
</tr>
<tr>
<td>Female</td>
<td>423</td>
<td>42.3</td>
</tr>
<tr>
<td>Missing</td>
<td>66</td>
<td>6.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>35 years</td>
<td>n/a</td>
</tr>
<tr>
<td>Missing</td>
<td>173</td>
<td>17.3</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2. Asthma

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>921</td>
<td>46.0</td>
</tr>
<tr>
<td>Female</td>
<td>874</td>
<td>43.7</td>
</tr>
<tr>
<td>Missing</td>
<td>205</td>
<td>10.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>35 years</td>
<td>n/a</td>
</tr>
<tr>
<td>Missing</td>
<td>129</td>
<td>6.4</td>
</tr>
<tr>
<td>Total</td>
<td>2000</td>
<td>100</td>
</tr>
</tbody>
</table>

**Key Point:** Missing or low quality data can and will change the interpretation of a study.

**Question: Validity**

- Validity refers to the accuracy of the collected data.
- Sources of errors:
  - Respondent provides inaccurate information
  - Data recorded inaccurately when collected
  - Data entered inaccurately into database

**Key Point:** Missing or low quality data can and will change the interpretation of a study.
ANALYZE AND INTERPRET SURVEILLANCE DATA

**Analysis in Support of Objectives**
- Analysis of the data needs clear, measurable objectives.
- Surveillance objectives guide:
  - Types of data collected
  - Types of analyses performed

**Key Point:** Always remember who your audience is and keep your presentation simple.
Descriptive Epidemiology

- Data used to describe the distribution of a health condition or event in a community.
- Person – who?
- Place – where?
- Time – when?

Person

Who?
- Age
- Sex
- Marital status
- Occupation

Question: Person

Prevalence of obesity (body mass index ≥ 30 kg/m²) among men aged 18-59 years old, Brazil, 2006

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Prevalence (%)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-25</td>
<td>4.3</td>
<td>3.1, 6.5</td>
</tr>
<tr>
<td>26-32</td>
<td>10.9</td>
<td>8.3, 13.6</td>
</tr>
<tr>
<td>33-40</td>
<td>14.9</td>
<td>12.7, 17.1</td>
</tr>
<tr>
<td>41-45</td>
<td>14.1</td>
<td>12.6, 16.3</td>
</tr>
<tr>
<td>45-69</td>
<td>16.5</td>
<td>13.9, 19.2</td>
</tr>
</tbody>
</table>

Key Point: Data related to place mean that they are a factor because of geography of a population.
**Question: Time**


**Risk Factors**

- Risk factors commonly associated with NCDs
  - Alcohol consumption
  - Diet and nutrition
  - Genetics
  - Lack of physical activity
  - Tobacco use

**Key Point:** NCDs are non-communicable chronic diseases.

**Types of Risk Factors**

- Modifiable — can be changed by the individual
  - Lifestyle choices

- Nonmodifiable — cannot be changed by the individual
  - Age
  - Family history (hereditary)
  - Race or ethnicity
  - Sex
Question 1: Risk Factors

Adapted from Chronic Disease Risk Factors Among Participants in Medical Examination, by Selected Demographic Characteristics

<table>
<thead>
<tr>
<th>High Blood Pressure</th>
<th>Age Groups</th>
<th>18-34% (SE)</th>
<th>35-49% (SE)</th>
<th>50-54% (SE)</th>
<th>55+% (SE)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported</td>
<td>2.6 (2.09)</td>
<td>11.3 (1.87)</td>
<td>37.9 (4.65)</td>
<td>34.1 (2.82)</td>
<td>15.3 (1.50)</td>
<td></td>
</tr>
<tr>
<td>Measured</td>
<td>3.4 (2.39)</td>
<td>26.3 (3.58)</td>
<td>66.2 (5.48)</td>
<td>51.4 (5.52)</td>
<td>30.2 (1.83)</td>
<td></td>
</tr>
</tbody>
</table>


Key Point: In this chart, a relationship between high blood pressure and age group exists.

Question 2: Risk Factors

Percentage of deaths over age 30 caused by tobacco, 2004


Interpreting Modifiable Risk Factor Data

- Individuals choose to engage in modifiable factors.
- Recommendations often encourage people to change behavior and make better health-related decisions.
## Key Point:
The key to interpreting surveillance data accurately is to know the limitations of your data.

<table>
<thead>
<tr>
<th>Slide</th>
<th>Notes</th>
</tr>
</thead>
</table>

### Interpreting Non-modifiable Risk Factor Data Tips
Use caution and sensitivity
- People do not choose to have non-modifiable risk factors.
- Children, women, and people from certain races or ethnic groups may be vulnerable populations.
- Vulnerable populations may be stigmatized if data are not interpreted appropriately.

### Interpreting Non-modifiable Risk Factor Data
- Be aware of non-modifiable risk factors.
- Cannot recommend changes.
- Can recommend reducing modifiable risk factors among people with non-modifiable factors.
  - Example: Risk of obesity increases with age:
    - Cannot modify age
    - Can increase physical activity or improve diet among older people

### Limitations of Surveillance Data
- Underreporting of cases
- Not representative of entire population
- Changes in case definition over time
**Underreporting**
Failure to report a health condition or vital event, as required by law, to proper officials
- Due to individuals being unaware of their responsibility to report
- Common in passive surveillance systems
  - Notifiable disease reporting systems
  - Vital events registration
  - Morbidity registries

**Example of Underreporting**
Can you share an example of underreporting that you may have experienced in your work?

**Representativeness**
- Definition: How accurately data reflect the occurrence and distribution of a disease in a population
- Affected by
  - Exclusion of particular subpopulations
  - Changes in reporting practices
  - Differences in reporting practices

**Key Point**: Surveillance data should be representative of the true distribution of a disease, so that health officials can effectively reduce the disease.
Key Point: If different case definitions are used for many studies (without acknowledgment), then trends in data may be interpreted incorrectly.
Ensuring Consistent Case Definitions

- International Classification of Diseases (ICD) codes
- International set of criteria used to classify health conditions and deaths
- Used for clinical, epidemiological, and managerial purposes
- Permits comparison of calculated morbidity and mortality between countries
- Periodic revisions to incorporate new health conditions as well as advances in science and technology
Instructions:
1. You will work in small groups to complete the assessment.
2. Select a member of your group to record your responses.
3. There are four parts to this assessment. Groups will have 75 minutes to complete the assessment.
4. At the end of the assessment, one member from the group will share your findings and recommendations with the class. (25 minutes)

Part 1. Background (10 minutes)
For this skill assessment, you will be examining CVD mortality and obesity.
1. If the objective of a surveillance system is to characterize and monitor trends in obesity in the US, what categories, and specific data within each category, should be collected?

2. In general terms, how would you present the data you collected in Question 1?
Part 2. Select Heart Disease-specific Mortality Counts and Rates in the US (20 minutes)

3. Using Figure 1, describe the trend observed in CVD-related mortality among males and females from 1979-2007.

**Figure 1.** CVD mortality trends for males and females (United States: 1979-2007.) The overall comparability for cardiovascular disease between the ICD 9th revision (1979-1998) and International Classification of Diseases, 10th revision (1999-2007) is 0.9962. No comparability ratios were applied. Source: National Center for Health Statistics.

4. Figure 1 summarizes the number of CVD deaths in males and females in the US. Why would it be better if Figure 1 showed rates of CVD mortality instead of just the number of CVD deaths?

5. In the title of Figure 1, there is a description about the change in the ICD 9th revision to the 10th revision. Does the change in the case definition affect your answer to Question 1? Why or why not?
Figure 2. US map of age-adjusted death rates for stroke by state (including the District of Columbia) in 2007.

6. What observations do you have about Figure 2 concerning the death rates of stroke in the US?
Part 3. Risk Factor: Self-reported Obesity in the US (25 minutes)


<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (N = 405,102)</th>
<th>Men (n = 158,455)</th>
<th>Women (n = 246,647)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (95% CI)</td>
<td>% (95% CI)</td>
<td>% (95% CI)</td>
</tr>
<tr>
<td>Total</td>
<td>26.7</td>
<td>27.4</td>
<td>26.0</td>
</tr>
<tr>
<td>Age group (yrs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–29</td>
<td>20.3</td>
<td>20.1</td>
<td>20.6</td>
</tr>
<tr>
<td>30–39</td>
<td>27.8</td>
<td>29.4</td>
<td>26.2</td>
</tr>
<tr>
<td>40–49</td>
<td>29.4</td>
<td>31.0</td>
<td>27.8</td>
</tr>
<tr>
<td>50–59</td>
<td>31.1</td>
<td>31.9</td>
<td>30.3</td>
</tr>
<tr>
<td>60–69</td>
<td>30.9</td>
<td>30.4</td>
<td>31.3</td>
</tr>
<tr>
<td>≥70</td>
<td>20.5</td>
<td>19.8</td>
<td>21.0</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>25.2</td>
<td>27.1</td>
<td>23.3</td>
</tr>
<tr>
<td>Black, non-Hispanic</td>
<td>36.8</td>
<td>30.9</td>
<td>41.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>30.7</td>
<td>30.6</td>
<td>30.8</td>
</tr>
<tr>
<td>Other race</td>
<td>16.7</td>
<td>16.9</td>
<td>16.5</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school graduate</td>
<td>32.9</td>
<td>29.6</td>
<td>36.4</td>
</tr>
<tr>
<td>High school graduate</td>
<td>29.5</td>
<td>29.5</td>
<td>29.5</td>
</tr>
<tr>
<td>Some college</td>
<td>29.1</td>
<td>30.6</td>
<td>27.9</td>
</tr>
<tr>
<td>College graduate</td>
<td>20.8</td>
<td>22.9</td>
<td>18.6</td>
</tr>
<tr>
<td>Census region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>24.3</td>
<td>25.2</td>
<td>23.4</td>
</tr>
<tr>
<td>Midwest</td>
<td>28.2</td>
<td>29.2</td>
<td>27.2</td>
</tr>
<tr>
<td>South</td>
<td>28.4</td>
<td>28.8</td>
<td>28.1</td>
</tr>
<tr>
<td>West</td>
<td>24.4</td>
<td>25.1</td>
<td>23.7</td>
</tr>
</tbody>
</table>

* Body mass index (BMI) ≥30.0; BMI was calculated from self-reported weight and height (weight [kg] / height [m]^2).
† Confidence interval.
§ Additional information available at http://www.census.gov.

7. Look at the different risk factors for obesity in Table 1.
   a. Which are modifiable risk factors?
   b. Which are non-modifiable risk factors?

8. Using Table 1, describe the prevalence of obesity in the US by person and place. Which groups have the highest prevalence of obesity in the US?

9. Using Figure 3, describe the time trend in obesity prevalence from 2000-2009.
10. BMI was calculated based on self-reported weight and height using the formula:

\[
\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height (m)} \times \text{Height (m)}}
\]

How do you think the prevalence would be affected if survey respondents said they weighed 2 kg less than their true weight?

**Part 4. Recommendations for Interventions (20 minutes)**

Using the information from Parts 1-3, consult your findings and use them to provide specific recommendations on a national program that will be designed to continue reducing CVD-specific mortality rates in the US by targeting the prevalence of obesity in the US. The recommendations should include priorities for specific geographic areas and subpopulations. For example, for which group should an intervention be targeted? *Note: Assume that height and weight were reported accurately.*
Description of Data Used in this Skill Assessment:
The CVD-related deaths were calculated using data from the National Center for Health Statistics and the United States Census Bureau. Mortality data are based on the underlying cause of death as reported on death certificates filed with the vital statistics office and compiled by the National Center for Health Statistics. Data on obesity prevalence in the US was obtained through the Behavioral Risk Factor Surveillance System, which is a national telephone survey that collects information on behaviors and health conditions, particularly non-communicable diseases. Body mass index (BMI), used to classify individuals as overweight or obese, was calculated from self-reported weight and height.

References: